

Data Stewardship and the Role of National Statistical Offices in the New Data Ecosystem: a brief

Key Messages

The Conference of European Statisticians (CES) Task Force **defines data stewardship** as *ensuring the ethical and responsible creation, collection, management, use and reuse of data so that they are used for public good and benefit the full community of data users.*

Data stewardship is necessary to maximize the value of data assets.

- Stewardship enables data policy implementation and the treatment of data as a strategic asset.
- Stewardship promotes sharing and reuse of data assets subject to the right ethical and cultural conditions, including data protection and (statistical) confidentiality, thereby maximizing their value.
- Stewardship promotes open data, interoperability and ethical and culturally appropriate use of data, enhancing the timeliness and efficacy of decision making and enabling data-driven public service delivery.

Data stewardship has two main scopes: intra-agency and system wide.

- Internal data stewardship can help consistently promote best practices in working with data within an agency, including improving data quality and metadata-based production.
- System-wide data stewardship requires cooperation and coordination across agencies which goes beyond various data domains. Achieving the desired availability and interoperability of data is a multi-party effort.

NSOs have inherent and unique expertise to lead data stewardship in the National Statistical System and to take on data stewardship responsibilities across the national data ecosystem.

- In their role of producing quality statistics, NSOs have always been engaged in data stewardship. This unique experience puts NSOs in a position to leverage its experience and functional expertise to take on new tasks and expanded roles related to data stewardship.
- The data stewardship role of the NSO can cover responsibilities in system-wide data stewardship related to public sector data management, depending on the national context.

Data stewardship improves data quality and relevance of NSOs and the National Statistical System.

- Data stewardship can support the delivery of high-quality data in all countries. Data stewardship brings added value to NSOs and the NSS, factoring for the national context affecting its implementation.

Data stewardship strengthens public trust in official statistics and in data management across the public sector

- Provision of data-driven public goods and services requires sound data governance, public trust and data reliability.
- Data stewardship supports the Fundamental Principles of Official Statistics which enable privacy, high quality and public trust.

Data stewardship needs governance, capabilities and resources for successful implementation.

- Data stewardship goals should be embedded in data governance - *a system of decision rights and accountabilities for the management of data, and the resulting laws, regulations, policies and frameworks that provide enforcement.*
- Data stewardship requires coordinated investments and capabilities, in transparent dialogue with the public and the budget decision makers to ensure progress and needed investments.
- A maturity model can be used for self-evaluating the gaps between a current state and desired end-state of data stewardship. This allows the definition of appropriate governance structures and planning investments.

Summary of *Data Stewardship and the Role of National Statistical Offices in the New Data Ecosystem*¹

CH1 Intro

The rapidly expanding and evolving landscape of data production and sharing is an opportunity for national statistical offices (NSOs) to reconsider their position and role in the national data ecosystem. A concept often used in association with this changing role is 'data stewardship'. As the transformation is multifaceted, there are different views of how the NSO role can and should change and what data stewardship means in this context.

To respond to these questions, in 2021 the Conference of European Statisticians (CES) set up a Task Force to define 'data stewardship' and related terms, and develop a set of possible responsibilities of a data steward and considerations for shaping the NSOs' role in this area. *Data Stewardship and the Role of National Statistical Offices in the New Data Ecosystem* is the result of the work of the Task Force, examining data stewardship through the lens of NSOs.

The work of this Task Force has focused on defining and explaining the foundational concepts and frameworks relevant to data stewardship. More concrete guidance and recommendations could be developed as a follow-up to this work. Suggestions for possible further steps to support NSOs in implementing a data stewardship role are summarized in full report Section 1.4 'Future work'.

The report aims to serve as a guide to statistical offices regardless of data or structural maturity. It will help offices to decide whether they should extend their role as public sector data stewards, as well as provide inspiration and ideas for those who are already performing this function. Public data holders will also gain a deeper understanding of the advantages and services provided by NSOs in the future data ecosystem.

CH2 Context

The digitalisation of society and economy offers both opportunities and challenges. There is huge potential for new types of data services, more timely and granular data, and new insights by linking data from different sectors, sources and topics. However, there are also considerable risks: increasing digital divide, heightened need for digital security and protection of privacy, and possible unethical use of data. Further, public agencies will have to address issues of data access, interoperability and standardization, data duplication and redundancy, and the costs of linking and sharing data.

To overcome the risks requires governmental policies with foresight, focusing on strategy, culture, ethics, roles and capabilities to support an effective data ecosystem. For this to work at an all-of-government level, there must be coordination, accepted norms, standards and efficient implementation.

When data becomes increasingly digitized, the desire to integrate data sets from different organizations for evidence-based decision-making increases. The existing data should be used efficiently, and no data should be collected if another public sector organization has already collected it (the 'once-only principle'). The path to better use of data is through increasing the interoperability of data and the use of common technical interfaces, facilitated by a public sector-wide alignment. To achieve interoperability, common guidelines, structures, metadata definitions, identifiers, quality frameworks, etc. should be in place. The statistical community has addressed significant parts of this problem decades ago by implementing international concepts, classifications and methods to promote the consistency and efficiency of statistical systems.

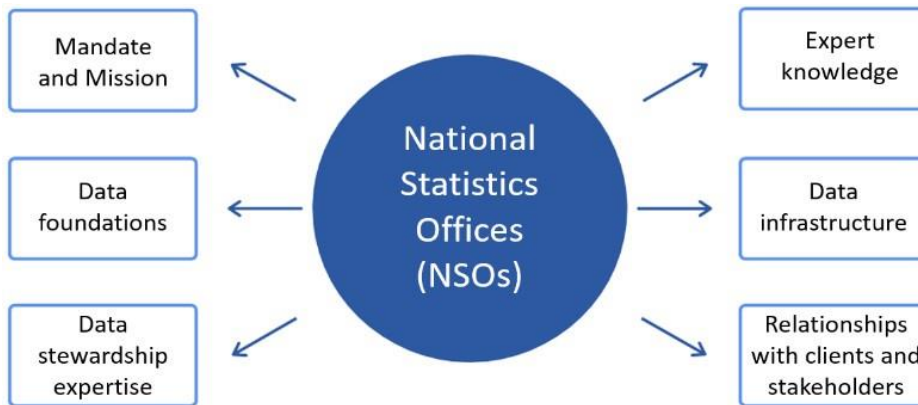
National Statistical Offices face the challenge – and the opportunity – to place themselves as key players in the data ecosystem given their well-grounded expertise in data management, access and sharing practices, as well as their key role as producers of statistics and indicators. Figure 1 provides a summary of capabilities of NSOs that give them a good basis to fulfil the data stewardship role.

The body of knowledge within NSOs can be used to promote interoperability and the use of standard classifications, metadata structures and quality frameworks in the broader NSS, among holders of administrative data, and among other institutions and organizations within the public sector. This expertise

¹ ECE/CES/2023/2. The blue boxes on the left refer to the Chapters in the full document.

is augmented by relationships with data holders who have compatible data and skillsets to NSOs (e.g. mapping and environmental agencies, central banks, register owners), which can be organized through partnerships.

Figure 1
Capabilities of NSOs to take on responsibilities in the data stewardship landscape



The degree of implementation of a leadership role in data stewardship will vary depending on national context, such as legal frameworks, digitalisation, the extent of cooperation achieved within the public sector, and public acceptance. There are risks either way. If NSOs fail to participate in the data ecosystem transformation, they risk losing relevance. Parallel data stewardship structures may be set up leading to further fragmentation of public data holdings with increasing burden and costs. However, there is also the risk of an NSO taking on a too ambitious data stewardship role and not being able to deliver expected performance that may interfere with its core operations and lead to reputational losses or decreased data quality. Also, the national cyber security provision which is classically not task of NSO will become even more important but difficult or impossible to manage for NSO in system wide scope.

When successfully carried out, NSOs will benefit by having secured an integral role into the future that will provide them with better quality data, access to more data sources allowing to produce more timely, relevant and disaggregated data; and increased possibilities for integrating data from different sources.

Maximizing the value of data assets in a data ecosystem requires common rules and standards as well as institutional setup. **Data governance** can be defined as a system of decision rights and accountabilities for the management of data, and the resulting regulations, policies and frameworks that provide enforcement.

Data stewardship implements the policies, standards and principles outlined by data governance. **Data stewardship** is defined as *ensuring the ethical and responsible creation, collection, management, use and reuse of data so that they are used for public good and benefit the full community of data users*. Data stewards manage and coordinate the interactions in the system, operating in service of - rather than in control of - the data ecosystem. In the context of NSOs, the definition emphasizes that data steward manages data on behalf and for the benefit of the whole society.

Data stewardship is applicable at all scales, from the national to data system level (system-wide stewardship), to the organization or agency level (intra-agency stewardship), to the individual or dataset. For the possible coverage of the NSO's stewardship role in the national data ecosystem, see Figure 2.

Figure 2
Possible coverage of NSO's data stewardship role



When shaping their role in data stewardship, NSOs must take into account their institutional position within the public administration. We can distinguish three main governance models, the two extremes being a fully centralized and a fully distributed model. The most widely used in practice and best suited for an all-of-government approach is the federated hybrid model. It has a centralized structure that oversees the

public sector's data environment but allows bottom-up input, enabling participation from individual agencies. The centralized structure provides a framework, tools and best practices for the agencies to follow but leaves the agencies autonomy to manage agency-specific data in line with their respective mandates. A data steward for the whole-of-government, and data stewards in individual agencies are required for optimal implementation of this data governance system.

The general definition of data stewardship covers a wide range of roles and tasks that data stewardship can entail in practice (see a summary in figure 3). These can include being involved in discussions on public data strategies, offering advice and guidance, providing methods and tools that other agencies can use, or taking responsibility as data stewards of (some of) the public data holdings. An important change compared to the role of NSOs as coordinators of NSS is that the stewardship tasks extend outside the statistical system, possibly even outside the public sector.

Figure 3

Summary of system-wide data stewardship work areas

<p>Data stewardship as a public sector goal</p> <ul style="list-style-type: none"> • Contribute to developing and promoting data strategies, policy and principles and build trust and confidence in the system <ul style="list-style-type: none"> ○ national /public sector data strategies ○ once-only, open data, FAIR, CARE principles
<p>Coordination and partnerships</p> <ul style="list-style-type: none"> • Coordinate standardisation and harmonisation process, support interoperability • Partnerships to improve capabilities, develop new and improve existing products <ul style="list-style-type: none"> ○ Data sharing arrangements ○ Partnership and licensing agreements ○ Guidance to other agencies
<p>Quality assurance and assistance</p> <ul style="list-style-type: none"> • Assess data quality in the National Statistical System and promote and provide expertise in data quality to other public data holders • Develop and promote quality frameworks (e.g. for administrative registers, privately held data) • Quality certifications and audit
<p>Data access</p> <ul style="list-style-type: none"> • Provide research and information services <ul style="list-style-type: none"> ○ Access to microdata, taking into account data protection and confidentiality ○ Surveys for other government agencies ○ Information services and analytical support to users (incl. policy makers) • Build data platforms and dashboards (for data sharing and dissemination) <ul style="list-style-type: none"> ○ Bring together data from different areas (e.g. Data Centres) ○ Links with geospatial data, geospatial visualization • Access privately held data sources
<p>Methods, tools and capabilities</p> <ul style="list-style-type: none"> • Standards and classifications • Infrastructure (e.g. for data integration, access to individual /micro data) <ul style="list-style-type: none"> ○ Data lake ○ (Meta)data catalogue ○ Common identifiers • Data integration • Data protection • Statistical methods, data science and machine learning methods • Common approach to data handling • Helping other agencies to build their skills, processes, tools and services • Improving data literacy

Statistical offices should be involved in the discussion on public sector data strategies, and aim to play an integral part in elaborating standards, quality frameworks, and other enabling instruments, given their expertise and the impact these enablers have on official statistics. In cases where the data stewardship role for the whole government sector lies with another institution, the NSO should extensively cooperate with the public sector Data Steward.

The responsibilities of a data steward within an NSO may be categorized by networking activities (both internal and external), technical responsibilities related to data management, and ethical and legal responsibilities. Within an NSO, a data steward would be at minimum responsible for:

- Data description (e.g. data catalogue, data dictionary)
- Metadata quality (the responsibility of the data lies with the data holder)
- Data life cycle management
- Data ethics (providing a link to the Ethics Committee)
- Data security, protection and confidentiality
- Data audits (monitoring the use of data).

Measuring data stewardship performance is relatively new territory, so this report also proposes a data stewardship maturity assessment model based on an example implemented in New Zealand. A maturity model can help embed important ideas and values into the understanding of what constitutes sound data stewardship and, by extension, into organisational structure and practice, to steer agencies in the right direction. A data stewardship maturity model can help creating a picture of what 'good' looks like, and help initiate change leading to concrete improvements. That knowledge will prove invaluable for an NSO's ability to operate effectively in a national data ecosystem leadership role.

Data stewardship can represent a challenging concept to understand. For its effective communication, it can be recommended to focus on describing the roles that a data steward can play in the data ecosystem, describe the challenges that good data stewardship can address, and how consistent, systematic and sustainable data stewardship can lead to a constantly developing data ecosystem creating public value beyond official statistics. The communication of data stewardship should be as simple as possible, starting with essential information only, and building in more detailed information over time. A practical solution to communicating data stewardship effectively can be to focus on what the purpose of data stewardship is, who performs the related activities and what these are. The communication can also address new products and services that data stewardship engenders and inspires.

A common understanding of the terms related to data stewardship is needed. A glossary (coordinated with the global Working Group on Data Stewardship) is provided in Annex 1. Annex 2 outlines how data stewardship relates to the Fundamental Principles of Official Statistics. Some other principles that are relevant and can be helpful when working on data stewardship are provided in Annex 3. A separate document as Addendum 2 to the report provides examples of how NSOs in different countries are pursuing their role as data stewards, and the work that Eurostat and OECD have undertaken in this area.